

Archiving For All:  
Working Towards Inclusive Digitization Standards  
Michelle J. Krasowski  
Internet Archive

### Abstract

At the Internet Archive we have been working towards the goal of Building Libraries Together, encouraging our users to archive and upload content from their locations. We hope to create a collection that can be used for research and discovery, representative of a diverse global community that has an interest in preserving and sharing the aspects of their cultural landscape with other users of different backgrounds, interests, and geographic locations.

As we develop the tools and instruction to engage our community in archival practices, it is important for us to consider the benefits and drawbacks of asking our community to adhere to professional archival standards. The limited access to higher cost equipment or professional archiving companies may be a deciding factor in what information gets preserved and passed along to current and future generations. At the same time, it is important to do what we can to make sure the best possible standards are achieved. How do we strike a balance to make sure communities with vastly different resources have a chance to participate and preserve? In what ways can we support them?

By highlighting the contributors that have worked with the Internet Archive to build robust and valuable collections, and by promoting solutions that share the information made available by preservationists, archivists, and institutions working in the field, we will identify ways that our community can help create and disseminate tools and information to help people preserve content in a way that is geographically, financially, and technologically possible for them.

## Archiving For All:

### Working Towards Inclusive Digitization Standards

As those in the profession know, the world of archives is by no means homogenous. With such a wide range of experiences in the world, and the different media with which these narratives are built, each archival collection is unique in its own way. The conditions under which archives reside are no exception to this variety. While there is a rich network of archives in the traditional sense – those in an institution's care, professionally preserved and curated with scope and available storage space in mind - there are also numerous collections that exist outside of this model that have their own important value. These other collections, with their unique purpose and potential to researchers and their communities, require equally unique solutions to help protect them and convert them to formats that will make them widely available.

The Internet Archive is a non-profit digital library that hosts archived materials shared from many different sources, each with their own standards and approaches to digitization of formats. The collections uploaded by our users range from files created following the highest standard of professional archival formats to lower quality digital captures that do not meet the guidelines set forth in the best practices of organizations including Library of Congress and the Federal Agencies Digitization Guidelines Initiative (FAGDI). We believe that achieving our goals of building a collection that is representative of a diverse global community requires a flexible approach to these standards; otherwise the stories of communities with greater economic and technical limitations are at risk of being excluded.

Earlier this year, the Internet Archive received a Knight Foundation grant for our Building Libraries Together project to develop the new version of our website, which will include tools designed to help our users add content more easily and to create communities around their collections. Along with these tools, there will be a need to provide solutions and documentation to help people prepare their artifacts for upload. Thankfully, there is a wealth of information already available as a result of the hard work of dedicated people in the Internet Archive community and the wider archival profession. With the right approach, we can work together towards increasingly inclusive solutions that help promote collaboration, education, and sharing of resources to support archiving for all.

### **Internet Archive as Research Institution**

One of the primary purposes of the Internet Archive is to offer permanent access for researchers, historians, and scholars to collections that exist in digital format. As of the end of April 2015 our collections contain 7.9 million texts, 2.4 million audio recordings, 964 thousand images, 1.9 million moving images, 102 thousand software programs, and 456 billion archived web pages (Internet Archive, 2015). As an online library we foster education and scholarship in the digital world, which opens up exciting new possibilities for access to a broader range of materials that is not limited by geographic proximity (Internet Archive: About IA, n.d.).

The growing world of online access to free and open resources that support education and research can be seen in the open educational resources movements, including the work of the Open Education Consortium ([oeconsortium.org](http://oeconsortium.org)) and the emergence of Massive Open Online Courses, or MOOCs. Similar educational models have emerged in the past making use of the new and existing technologies of the time, and the

perpetuity of this ideal today demonstrates that open and convenient access to educational resources is an ongoing need (Massive open online course – Wikipedia, 2015).

With the definitions of higher education and research institutions expanding in the online realm, a more inclusive view of libraries and archives can follow to open the way for more diverse collections. The limitations of physical space and access are being overcome by the ability to create an online collection. A virtual environment now provides room for people who are experts in their fields of interest or study and have their own privately curated collections, enthusiasts and hobbyists who have a passion for a particular subject area, and creators of the original content housed in their own collections. They all have the equivalent of shelf space to share their scholarly efforts on the Internet Archive and to make it available to the larger community who can benefit from their work in these collections.

### **Collections in Peril, Users in Need**

Deterioration of magnetic media is a concern of collection holders everywhere. Temperature and humidity regulation are two controllable factors that can help media have the longest life possible, and continuing improvements in hard disk storage capabilities means that the associated hardware cost of creating a digitally archived version of a work is on the decline (although equipment and labor costs continue to be a major factor). For institutional archives, digitization efforts will naturally focus on items that are already a part of their curated collection, meaning that the materials have experienced a certain degree of maintenance. Their risk of deterioration is an impetus to begin the task of digitization, and the same can be said for other collections that are not in such stable environments.

All across the world, there are collections with valuable content that have not existed in conditions conducive to preservation. They may consist of VHS cassettes of broadcast news stories collected on a certain topic by a professor for classroom use, audiocassettes of interviews with members of a community that gives valuable historical information, or recordings on formats soon to be obsolete and unplayable. They may be located in crowded apartments, attics that experience temperature and humidity fluctuations, or storage sheds that are open to the elements and inviting to mold, pests, and other damage. They may be in a region that is at high risk of natural disaster or flooding due to climate change, or in an area where residents are being displaced from their homes on a frequent basis due to political conflict, current economic conditions, and gentrification. When these circumstances uproot people from their homes and their lives, there is a significant potential that these materials will be lost since many times the people who are affected don't have the ability to keep all of their possessions and bring them along. These are the collections that are most at-risk, which tell the regional and folk history of society and provide insight into diverse communities whose stories will disappear without the support of the archival community.

Many people living in these circumstances also do not have the financial resources to hire a preservationist to digitally archive their content. It is important for members of the profession to continue advocating for funding from the government to preserve collections of historical value, and to find methods of digitization that are time and cost effective that can be replicated in many communities. Establishing a network of community digitization centers, either as a part of existing institutions such as public libraries or universities, or in new locations, would be a way to make sure resources are made

available to the widest range of needs. In particular, this would assure that funding from the government and from grants would be allocated in a way that could benefit smaller organizations that may have difficulty qualifying for grants on their own. User-friendly equipment and workflows could provide an opportunity for more collections to be transferred and shared, and creating positions for trained members of the archival field to educate interested parties within communities on simplified processes could lead to digitization efforts that become self-sufficient and spread to others who are interested in taking part. Thankfully, this is a need that has been identified in the past and still recognized today, with many inspiring projects to motivate us.

### **Building Collections**

The scope of collections housed on the Internet Archive has benefited greatly from individuals who understand the possibilities provided by access to free storage and the ability to use the site as a platform to share resources with the community, and who have devised solutions to make digitization possible for people with limitations. Two major content contributors who have brought a significant amount of valuable material to the archive through their own passionate work are Skip Elsheimer and John Hauser.

#### **Skip Elsheimer – A/V Geeks**

<https://archive.org/details/avgeeks>

Skip is an important contributor to the Internet Archive, both through uploads to the A/V Geeks collection from his own archive of ephemeral films and his work on other collections. These include thousands of uploads from partner organizations and clients which increase our holdings of educational resources in a wide range of subjects, with particularly focused moving image collections in chemistry, mathematics, and computer

science, as well as archeology and anthropology. His work as a consultant, vendor, collaborator, and educator has made a profound impact on the success of many individuals and organizations that seek out available options for digitization from professional standards to low-budget, grassroots efforts.

### **John Hauser - Community Media Archive**

[https://archive.org/details/community\\_media](https://archive.org/details/community_media)

The Community Media Archive (CMA) is a collection of diverse local programming contributed by community access television productions from across the country, including channels serving a wide range of Public, Education, and Government (PEG) purposes (The Community Media Archive, 2009). Originally a partnership between the Internet Archive and Access Humboldt in late 2008, the CMA has grown to ingest video by over forty Access Centers thanks in large part to the efforts of John Hauser, Special Projects Manager at Access Humboldt, as well as his colleagues. (Access Humboldt – Wiki, 2015). John has been a regular presenter at Alliance of Community Media conferences since 2009, where he provides easily comprehensible instructions for creating collections with and uploading to the Internet Archive. In his presentations, John has been an advocate for more flexible digitization standards so that resources that are awaiting conversion are made accessible sooner rather than later (or not at all). He is constantly developing methods for centers without high bandwidth capabilities to enable them to store and submit their content to be uploaded offsite, and improving the metadata of the collections in the CMA for increased access and discovery.



**In-house A/V Digitization at the Internet Archive**

In addition to the independent contributions of material digitized offsite, the Internet Archive has limited in-house digitization capabilities for audiovisual material. While we do not meet professional standards, the individuals that come to us with donations for digitization and upload have still appreciated our services. Our equipment is selected and maintained by Sam Stoller, an engineer on our Petabox team. I have the pleasure of overseeing digitization projects, which includes devising workflows for different collections and training volunteers on processes and metadata entry for upload. We currently have equipment for digitizing CDs, DVDs, LPs, audiocassettes, VHS/SVHS, Beta, BetaCam SP, and UMatric. Visual materials are transferred from magnetic tape to DVD and then uploaded as an ISO file. Audiocassette signals are captured via the Creative Soundblaster ADC and recorded to Audacity via the line-in USB setting at a sample rate of 24bit/96000 Hz and saved as FLAC. None of the files digitized in-house are processed for noise reduction or signal improvement, but the files are made available for download to the collection donors should they wish to perform any post-processing.

While our workflows have room for improvement, two interesting collections are worth highlighting here to demonstrate that our capabilities can still be considered worthwhile for people in need of low-cost solutions. In 2014, the Internet Archive received a donation of VHS tapes from Dr. Michael Aldrich, cannabis scholar, medical marijuana activist and former curator of the Fitz Hugh Ludlow Memorial Library. The collections, recorded by Dr. Aldrich from Bay Area television stations between the years of 1986 and 2006, were focused on drug-related news stories of the time as well as the emerging AIDS epidemic. His gathering of the media portrayals of both legal and illegal substances as well

as sex, sexuality, and the harsh and hypocritical policies of the Reagan and Bush administrations of the 1980s provide valuable source material for scholars and activists that are researching the disastrous war on drugs or marginalization of the minority communities first impacted by HIV. Since many of the VHS tapes were compiled from earlier recordings, taped straight from broadcast television, and recorded at SLP, the quality is already low. To pay professional hourly charges to digitize this collection is probably not the best use of funds or high-quality equipment. However, since the content is the most vital part of this collection, our workflow was deemed an appropriate option. The collections, named The Dope Tapes and The AIDS Tapes, reside in our Ephemeral Films category on archive.org.

Another collection that I have recently started digitizing was brought to us by Neil MacLean of the Ohlone Profiles Project (<http://ohloneprofiles.org/>). This organization has over 1,000 audiocassettes of the “Voices of Native Nations” program hosted by Mary Jean Robertson on local radio station KPOO since 1972, as well as poetry readings, tribunals, and HDV cassettes of Ohlone cultural presentations and gatherings in San Francisco from 2010 to 2014. They have applied for grant funding for digitization in the past but have come up against numerous obstacles in the requirements to qualify for professional digitization services. We are working closely to identify key parts of their collection to prioritize for digitization at the Internet Archive so that we can provide back-end storage to their WordPress site, where photographs of Ohlone cultural presentations and official documents from national and state parks, the city's planning department, Office of Human Rights, and Arts Commission, and private organizations are already featured. (N. MacLean, personal communication, April 23, 2015). This collection is particularly important for

political and activist reasons to make the case to the city of San Francisco that the Ohlone have cultural practices that deserve increased support and inclusion in the city's future. Neil will be presenting on our collaboration at the International Conference of Indigenous Archives, Libraries and Museums (ATALM 2015: <http://www.atalm.org/node/63>) to encourage other tribes to work with the Internet Archive as a platform for storage and information dissemination.

### **Inclusion Solutions: Past, Present, and Future**

There are many possibilities for working together to share resources that will be of benefit to both institutional archives and smaller organizations and individuals. There is already a wealth of information available as the result of hard work by members of the archival profession, and exciting projects are underway to provide resources for education and collaboration.

One of the traditional community resources for education and skill sharing is the public library. There have been a number of successful digitization labs launched across the country, including the Forsyth County Public Library system and Wake Forest University in North Carolina, The Hub at the Kalamazoo Public Library in Michigan, the Arlington Heights Memorial Library in Illinois, and the Shelby White & Leon Levy Info Commons at the Brooklyn Public Library in New York, to name just a few. Each of these locations has staff to help train and support the patrons who come in to use the equipment. Signups are often required, as well as adherence to a time limit, which would restrict the ability to use these stations for large-scale projects. Funding comes from private sources as well as from the individual states and organizations such as the Institute of Museum and Library Services (IMLS).

Other interesting digitization solutions include temporary pop-up stations that are tied to an organization's focused project, such as the "XFR STN" hosted by the New Museum in New York City from July 17 through September 8, 2013. This exhibition/lab allowed artists to schedule a three-hour appointment to receive advice on best practices and transfer, and to have their selected media migrated to digital format. Transfers were done at preservation grade, and files were uploaded to the Internet Archive (XFR STN, n.d.). By making the files part of the Internet Archive's collection, the New Museum was supporting the concept of "distribution is preservation" with the view that circulation is a mode of conservation (XFR STN Project, 2013, p.4). The XFR Collective is still going strong and providing training and support for other people looking for help in the arena of digitization. You can visit them at <https://xfrcollective.wordpress.com/> to access their excellent Resources page and keep current on their activities in the Goings-on section of their website.

The Bay Area Video Coalition, or BAVC (<http://bavc.org/>) is another fantastic organization that is doing important work to make information and services available to people undergoing digitization efforts. They have made preservation at the highest standards with significantly reduced costs possible for organizations that would otherwise not be able to afford this quality of work through the Preservation Access Program, which subsidizes their work with grant support from the National Endowment for the Arts and Getty Research Center. They have also developed tools for people undertaking digitization projects. Their free QCTools software is specifically designed for archival video capture in order to improve the efficiency of the process and eliminate the chance of unintended reformatting by production and editing software (Quality Control Tools, n.d.). The

Audio/Visual Artifacts Atlas (AVAA), a resource developed in partnership with Stanford University and NYU, provides helpful reference videos and glossaries that enable the identification of artifacts and errors in analog and digital media, along with information on the correction of these errors if it is possible. The AVAA is a community project and users are encouraged to contribute additional materials (AV Artifact Atlas, 2015).

Finally, at the end of April 2015, BAVC announced a new resource called AV Compass, set for release at the end of June. Developed with a grant from the Andrew W. Mellon Foundation and with recommendations made by the California Audiovisual Preservation Program (CAVPP) and KQED, this will provide “a free, web-based resource designed to empower those who lack training or confidence in the skills needed to preserve their audiovisual materials to begin the process” (BAVC, personal communication, April 24, 2015). With documents, videos, and inventory tools to help people find a solution to their preservation needs, this is a platform with extremely exciting possibilities.

This is an excellent time for anyone who has visions of how such a platform could be utilized to its fullest potential to contact the development team at BAVC with ideas and feedback. My suggestions include creating sections for equipment donations and sharing of equipment between institutions; a have/want marketplace for equipment and supplies so that members of the community are not competing for equipment and driving prices unnecessarily higher in online auctions; a regional schedule of training classes for equipment maintenance and repair; a “boneyard” where people can offer up machines in need of fixing or service, or to be used for parts; an equipment manual library; a roundtable to strategize ways to petition for more funding from the government for these important projects; and a forum where users can connect, communicate, and collaborate. It is also

important that we think ahead to the future and what will be required to access material stored on hard drives, not only in terms of formats but also in terms of access to a power supply and the life expectancy of hardware. I'm sure the team at BAVC already has many of these features in mind, and I look forward to spreading the word about this resource far and wide, as well as reaping the benefits to improve our capacities at the Internet Archive. On behalf of my institution's user community, I would like to express my appreciation to Lauren O'Connor, Preservation Resources Fellow at BAVC who has already put so much effort into this project, and I encourage as many people as possible within the archival profession to contribute to this effort.

The people and projects mentioned in this paper are but a few of the amazing initiatives I have come across in my research and exploration of the archival efforts for preservation happening in the world. I am grateful to have had the opportunity to share my enthusiasm for their work, and for the opportunity to learn from countless others who are making information accessible. Thank you, every one.

## References

ABOUT << Ohlone Profiles. (n.d.). In *Ohlone Profiles*. Retrieved April 25<sup>th</sup>, 2015 from

<http://ohloneprofiles.org/about-2/>

A/V Geeks. (2005). In *Internet Archive*. Retrieved April 25<sup>th</sup>, 2015 from

<https://archive.org/details/avgeeks&tab=about>

AV Artifact Atlas. (2015). In *AV Artifact Atlas*. Retrieved April 25<sup>th</sup>, 2015 from

[http://avaa.bavc.org/artifactatlas/index.php/A/V\\_Artifact\\_Atlas](http://avaa.bavc.org/artifactatlas/index.php/A/V_Artifact_Atlas)

Community media archive – Access Humboldt. (2015).

Retrieved April 25<sup>th</sup>, 2015 from the Access Humboldt Wiki:

[http://accesshumboldt.net/wiki/index.php?title=Community\\_media\\_archive](http://accesshumboldt.net/wiki/index.php?title=Community_media_archive)

Digitization | North Carolina Room – Forsyth County Public Library. (2010). In *North*

*Carolina Room – Forsyth County Public Library | Genealogy, local history, culture, and government*. Retrieved April 25<sup>th</sup>, 2015 from

<https://northcarolinaroom.wordpress.com/services/digitization-center/>

Digitization Assistance for Community Media. (2015). In *Digitization Assistance for*

*Community Media*. Retrieved April 25<sup>th</sup>, 2015 from

<https://xfrcollective.wordpress.com/>

Digitize your VHS and Videocassettes. (2015). In *Arlington Heights Memorial Library*.

Retrieved April 25<sup>th</sup>, 2015 from

<http://www.ahml.info/content/digitize-your-vhs-and-video-cassettes-10>

Internet Archive: About IA. (n.d.). In *Internet Archive*. Retrieved April 25<sup>th</sup>, 2015, from

<https://archive.org/about/>

Internet Archive: Digital Library of Free Books, Music, Movies, & Wayback Machine. (n.d.).

In *Internet Archive*. Retrieved April 25<sup>th</sup>, 2015, from <https://archive.org/>

Kalamazoo Public Library. (2014). In *Kalamazoo Public Library*. Retrieved April 25<sup>th</sup>, 2015

from <http://www.kpl.gov/hub/>

Preservation Services. (2013). In *Bay Area Video Coalition*. Retrieved April 25, 2015 from

<http://bavc.org/preservation/services/history>

Massive online open course. (2015). In *Wikipedia*. Retrieved April 25<sup>th</sup>, 2015 from

[https://en.wikipedia.org/wiki/Massive\\_Open\\_Online\\_Course](https://en.wikipedia.org/wiki/Massive_Open_Online_Course)

Quality Control Tools for Video Preservation. (2013). In *Bay Area Video Coalition*.

Retrieved April 25, 2015 from <http://bavc.org/qctools>

Shelby White and Leon Levy Information Commons. (2015). In *Brooklyn Public Library*.

Retrieved April 25<sup>th</sup>, 2015 from

<http://www.bklynlibrary.org/locations/central/infocommons>

The Community Media Archive. (2009). In *Internet Archive*. Retrieved April 25<sup>th</sup>, 2015

from [https://archive.org/details/community\\_media&tab=about](https://archive.org/details/community_media&tab=about)

XFR STN at the New Museum | about. (2013). In *New Museum*. Retrieved April 25<sup>th</sup>, 2015

from <http://xfrstn.newmuseum.org/about/>

XFR STN Project. (2013). In *Internet Archive*. Retrieved April 25<sup>th</sup>, 2015 from

<https://archive.org/details/xfrstn&tab=about>